Putting Digital Libraries to Work for Education

Dr. Miriam J. Masullo
Thomas J. Watson Research Center
IBM Research Division

Dr. Linda Tsantis
America Tomorrow, Inc.

OECD-KISDI Joint Conference on Information Infrastructure
Seoul, Korea
April 1995

Abstract

This paper reports on a series of activities covering the broad scope of digital libraries in education. It is intended as an overview to familiarize the audience with the issues by means of practical examples. It traces one pioneering effort to put digital libraries to work for education.

Introduction

Equity access to information is fundamental to the intellectual development of a nation. The ability to create new sources of information, and to make those available to anyone, anywhere is certainly a well-understood goal that all information infrastructure and superconnectivity plans have in common [1]. Digital libraries while not fully defined or well understood yet, are expected to help realize that goal. It is important to sort out the issues, separating those that we do understand from those that require exploration. This paper describes practical efforts at understanding and working on the issues and related technologies. The section on the Role of Digital Libraries in K-12 Education reports on a pioneer workshop by the same name. This workshop marked the first concerted effort to map digital library activities to the domain of K-12 education. The section that follows reports on a case study, EduPort [2], a demonstration of a vision for the role of digital libraries in education, also a first attempt at bringing the enabling technologies into the classroom. These sections deal with the topic of this paper in terms of two distinct and focused experiences. The next section deals with a different kind of experience, the goals and substance of one legislative effort. Policy is as crucial to the issues as technology, and perhaps more important to the outcomes.

The activities described in these sections, have some historical significance in that they are each first attempts in their class, at making appreciable progress and staking new frontiers of participation for our schools, in the national digital libraries initiatives. Against the background of these experiences, a section is dedicated to discussing the possible evolution of the information environments and infrastructures of the future. The Conclusion section summarizes current positions and invokes hopes for future work.

Background

By the end of this century large-scale digital libraries containing the collective legacy of human knowledge and information will be accessed in real-time, on-demand, over gigabit networks. The use of these technologies in education holds great promise for both students

and educators. Digital libraries will provide access to information currently available in books, videos, and other traditional forms, while digital museums will provide access to content that is not currently available in these media, and that is generally limited in accessibility, particularly to students. The potential to be able to interact with limitless sources of educational content, in real-time on-demand, both from classrooms and homes will change the way in which our teaching and learning environments are organized. And, curriculum, from the perspective of a methodology used to organize information in such a way that learning can be manageable, may have to be once more reformed to take advantage of these opportunities. Workshops and demonstration projects are extremely important because they give parents, educators, and decision makers the ability to explore and envision possibilities.

Workshop on the Role of Digital Libraries in K-12 Education

The 1993 Conference on Information and Knowledge Management, in Washington D.C. hosted a workshop on the Role of Digital Libraries in K-12 Education. This workshop was organized in collaboration with NASA and the North Central Regional Education Laboratory (NCREL). Three panels were assembled to deal with the broad scope of issues relevant to this topic in the areas of technology [3], application [4], and policy [5]. This section captures the collective thoughts and ideas of panelists as accurately possible.

Dr. Milt Halem of NASA Goddard Space Center set the stage for the work of the panels with an inspiring keynote address recognizing the difference between the currently evolving digital computing systems revolution and the forthcoming data revolution. Dr. Halem stressed that the technology for the first is the enabler for the other. Bringing these together is the grand challenge of our time. But, the impact of these revolutions is making little in-growth in our K-12 environments, and that has not yet been taken up as a challenge. Dr. Halem illustrated the meaning of this challenge by tracing one evolution of information turned into knowledge through a process that started in 1958. Effectively pointing out that we are almost half a century into the Information Age.

The impetus for the current initiatives revolving around the concept of digital libraries started with two decades of planning since 1958, followed by two major events associated with the history of NASA. President Kennedy kicked off NASA's Global Atmospheric Research Program in 1961 in his address to the United Nations, where he proposed an international collaborative effort of all the nations to intensify analysis and measurement of the global atmosphere from space and earth. The launching of NIMBUS-7 in 1978 was part of this effort. And, the data collection that resulted has implications for our society having dramatically changed our policies and programs since then.

The data collected has generated much knowledge that we now take for granted, but the data itself remains vastly unexplored revealing how little we know about worlds of information already in existence. Dr. Halem presented an array of formidable visual data, gathered from that program, representing: 14 years of monitoring the hole in the ozone layer over Antarctica, vegetation patterns on the ocean floor throughout the seasons, changes in snow fall accumulation over the North Pole, the relationship of warm water currents in the Pacific Ocean to weather patterns, the differences in the movement of clouds over the Northern Hemisphere and the Southern Hemisphere, and other examples.

The combinations of images and simulations that were produced are not only of scientific value but also bring about an appreciation for the aesthetics of information. Such data is a dynamic source of information, a visual information system. And, it forces us to deal with a completely different domain of information outside that which is perceived by the senses, since it would be too difficult or impossible to understand these phenomena in any other way. But, we don't have a natural model or system to guide us in abstracting information from such data. The task of interpreting data, and deriving knowledge from data at this level of detail is extremely difficult, yet we now take the results for granted, because they have been obtained by a few who have understood the value of data and know how to deal with it. The role of digital libraries in education begins with understanding how to interpret information, and that was well put forth by Dr. Halem's presentation. The time is over for making our students computer literate; the future demands that we make our students data literate.

The Applications Panel (schools libraries, museums)

Students seem to learn in ways different from the way in which their teachers learned. Therefore, regarding the digital libraries issues, for librarians and teachers the focus is more on knowledge vs. learning than on knowledge vs. information. Questions of quality of information, not quantity or availability are the concern of this group. What kinds of information we have (good or bad) is relative to the intellectual problem that we want to solve, therefore there is an issue of relativity associated with the quality and value of information. Assessment of this relative value of information issue will be magnified in the digital library paradigm.

Another important issue is that of mission. The mission of a library or museum is to curate and cultivate content. In a way the immense work that NASA has done to extract knowledge out of data is still the best model that we have for accomplishing that mission. In this new paradigm we may also find an amplification of the curatorial functions that so many organizations are now performing as a service to society. Also, we must consider that the classroom as a physical entity associated with learning, revolves in many ways around a concept of time, the time students spend with their teachers and other students. If this construct changes then schools will require restructuring and it is unknown what will this mean for the teachers, it is likely to mean more work in terms of management of activities and preparation of materials. Preparation of materials for "classroom" use, however, might not change dramatically in concept. This task may in fact be facilitated by the digital media, as might the extraction of knowledge. Term textbooks may become the tools of the teachers. Traditionally teachers have prepared materials by selecting from available media and combining it to create learning experiences. This activity could become more efficient and teachers more productive.

Examples of interdisciplinary experiences like the San Francisco Cardboard Kayaking, where students are asked to build and sail a kayak, may become the norm. These kinds of projects may be the best way to prepare students for dealing with the wealth of information that will become available on-demand. Interdisciplinary approaches to knowledge acquisition may be the key to make sense of information. As schools begin to connect with information servers, regional structures for organizing our education system may be considered. The concept of community information centers may start to make more sense, but this in turns may have profound implications to our national culture. Libraries for

example, (in their role as community information centers) encourage the kinds of programs and activities that are relevant and meaningful to the communities and the schools. It will be difficult for a digital library, whose functions are driven by equipment and data, to undertake that rile. What will happen when we remove the human factor form the equation?

The Technology Panel

Issues regarding technology are typically driven by the identification of problems within specific application (of technology) domains [6]. The identification of these problems regarding digital libraries turns out to be a daunting task when it comes to K-12 education. The technology panel could only begin to present the problems. Isolation of the issues will take a long time.

In order to begin to understand the situation we must consider that, as suggested by panel members, most of the technologies are being driven in their development by the larger more generic problems that apply to all domains. Applications drive the development of the technologies, and in this case applications are not exclusively coming from the education arena, although education is recognized as an important application area. Within the scope of the larger digital libraries activities, issues that have been identified by technologists are in the class of content volatility, that is, how often will data change in these environments, and information throughput, how many users will access the information, with what frequency, and at what level of concurrency.

These issues relate mostly to the communications platform and drive requirements in terms of representation and organization of information for interacting with it. How we represent and organize information for these environments is predicated on the mode of access [7]. These are significant consideration that will determine how to scale up the environments, and how they will integrate and interoperate. Dedicated testbeds will be necessary for exploration. Results obtained in general purpose testbeds are not likely to help with the K-12 problems.

Education testbeds for digital libraries will differ from other testbeds in some important ways. One specific goal here is to create networked environments where the users are also the providers of the information. But, with conceivably limitless sources of information and the incorporation of the larger population of users turned providers, how will educational choices be made? The concept of co-laboratories of information as opposed to repositories was presented as a possible solution. The problem is related to the relationship of information access to that of people access, and can be articulated as: how to include the vast source of educational resources associated with people access within a single infrastructure. Tools for managing these kinds of resources and collaborations across networked environments are needed, and must be developed with education in mind.

An example of the problem generated by networked environments where people become resources (in the same class with content), becomes evident when we think of how thousands of students will gain access to elite groups of very busy scientists. Information management issues (if we consider these scientists as sources of information) begin to gain complexity. That problem does not only drive requirements for tools to manage access to information of all sorts, but also indicates that there is a need to train our students to become inquiry literate as well. A structure of inquiry hierarchies merits consideration.

Also, the notion of working with ideas closer to environments that will work years from now is essential to successful implementation, if we are interested in building these to be of benefit to our students, this is, of benefit for the future [8]. And, if we are to overcome the current perception that there is an implementation bottleneck where the educational community is concerned, then it becomes essential that we put these tools in the hands of the teachers as an integral part of the solution.

The Policy Panel

Two key policy issues that capture the concern of the communities at this time seem to be in the areas of subsidy for the schools, and standards for the technology providers. In a sense standards mediate between the digital library users, in this case the schools, and the information providers, in this case, not excluding the content providers, but more specifically the technology providers. The cost associated with providing the digital technologies and the connectivity is an issue also related to standards. So it appears that the question of standards may have to be answered before any significant progress can be made. However the communities that will formulate and provide the technologies are divided in this issue. Universities and research laboratories prefer a process of chaos and natural selection in order to arrive at standards. Industry on the other hand, requires standards to help bring their products to market, particularly in this area where large investments are at stake.

The ownership of these digital libraries is likely to be in the hands of those who will develop them, and that is unlikely to be the government, or to be managed by the government. Government should not inhibit the development of infrastructure and digital libraries [9]. Policies however are formulated at government levels and go through tedious government processes. With this in mind, and with education in mind, what can be done to facilitate policy development to support the goals of the schools? The imperative is 1) to be as informed as possible on the issues, and 2) to speak out to those who are in a position to make change.

Beyond that, criteria for good policies lies in what we as a nation wish to get out of these revolutions in communications and information. Equity in access is the prime hope that expresses the informational needs of our students. In the current model access is free, when we come to a library, but information is not free, the library has to buy information. In the model of the future access is unlikely to be free, and information will continue not to be free. So what will happen to those who have limited access now?

It is necessary to reach a consensus among technology providers, information providers, government, and the public, on where we want these imminent changes to take us. In the scope not only of a democracy, but also of a civilized society, information must be made available to anyone who can benefit from it. These things must be written into legislation, and legislation must play a key role in sorting out all the issues, in all of the areas. There cannot be shortcuts or omissions. In the many issues were raised in this workshop, and the many more that need to be identified, lies the burden of building the future with technology.

Case Study: EduPort

EduPort is a demonstration project, designed to suggest and explore specific ways in which the digital libraries and communications networks of the future might be used for education. The project was first launched in the US during National Science and Technology Week in April of 1994, at Lincoln High School in Lincoln, Nebraska, where it is currently in use. In May of 1994 EduPort was demonstrated at the Challenge Nebraska Conference in Lincoln, Nebraska.

Overview of the Project

EduPort is made up of (among other things) a digital library containing newly digitized materials, approximately 95% of which is digital video. The digital library is housed at the University of Nebraska and accessed from a large-scale video server. A high-speed fiber optics line was installed to connect Lincoln High School to the video server. The content for EduPort comes from organizations in the US that have large inventories of media in the public domain, and that wish to make their content available to the public.

In the classroom the video is selected and controlled with a remote control. A single large electronic blackboard in the back of the classroom equipped with speakers is used to view the videos. The teacher is free to walk around the classroom and interact with the students and with the content at will. All the interactions are in real-time on-demand. That is, any video any minute. The teachers can stop and start the videos at any time. Teachers and students can interact with the materials at their pace, and link to, or switch to other materials depending on what the situation and interests are at the time, as opposed to having to depend on predetermined links built into hard-coded software by a software developer at some point in the past.

This new approach is possible because all the content is resident and accessed directly from the digital library at the university and no download is necessary. Based on this approach three scenarios were identified. The classroom scenario as described above, and the home scenario are identical with the only difference that the content in the home scenario would be accessed from home, by means of interactive television facilities. The home scenario in the context of EduPort is only currently implemented in the laboratory, but the technology has been demonstrated in association with the project. The development of the teacher work center scenario is underway, and would be used to simplify the task of creating and organizing content for EduPort, which is now labor intensive and requires special skills. Access to the teacher work center would be from a workstation or an interactive TV environment. There are no boundaries that separate what teachers and students would gain access to in these scenarios. The issue is point of access.

Organization of the Content

The digital content for EduPort was created from materials that were collected from organizations such as NASA, the Franklin D. Roosevelt Presidential Library, the Smithsonian, the National Gallery of Art, the Kennedy Center for the Performing Arts, and many others. The materials from these organizations that went into the creation of EduPort were not previously available in digital form. Thus a digital subset of these organizations now resides in a digital library at the University of Nebraska. Indeed the main reasons for

creating and using these digital libraries are: preservation and increased opportunities for access.

The materials are organized in a sort of video encyclopedia fashion to facilitate search and retrieval within a common pervasive well-understood mechanism. The organizing blocks are: Arts, English, Geography, History, Languages, Mathematics, and Science. Within those, the content is then organized into media blocks. Each media block contains a given combination of media objects, typically one or more video clips, images, or text (an example of text would be the digitized manuscript of an FDR speech). The idea is to provide a methodology for bringing these materials into the classroom within the relevant context of what is going on, in a systematic, organized manner.

The teachers, who made the selections based on existing curriculum, first combined Media objects to create media blocks. The educational context is thus created by the teacher, guided by existing curriculum, and modulated by the current needs and interests of the students. Each media block can contain media objects from one or more source or organization (content provider). And, in addition to the new materials that were digitized specially for EduPort, existing available sources of digital media can be added to the EduPort digital library, and made available to the teachers.
Among the many advantages that this system provides, is the instant access, remote in real-time. This has profound implications that greatly simplify deployment in the schools, since it reduces the hardware and software requirements for the schools, and paradoxically, the level of complexity that the schools have to deal with to leverage this advancement in technology. This also has favorable implications on cost factors.

System Components

The components of an EduPort-like environment are:

1. A video server: capable of supporting a given number of users measured in terms of simultaneous video streams; and a given quantity of materials measured in terms of storage.
2. A communications link: between the server and the clients capable of handling the transport traffic at the required speed for smooth flow of the media.
3. The players or receivers: that can decompress and play the media at the receiving end.
4. The content or digital library: of digital media, with appropriate user interfaces.

Specifications for implementation vary depending on the components selected or available in each case (e.g., what size server, what kind of communications link, etc.). The amount of content that can be made available is a function of the compression rate of the digital media, the amount of storage available, and the I/O throughout of the server. The number of users that can be supported depends also on the throughput of the server, as well as on the speed of the communications link.

A Legislative Effort

Changes to our laws do not happen spontaneously. Advocacy is required to motivate and push forward difficult changes. It is the task and responsibility of lawmakers to hear and

see the cases presented before them. A demonstration of EduPort was brought before the U.S. Senate Commerce, Science and Transportation Committee as it prepared to propose legislation that would affect and impact access and delivery of information for educational purposes. Several measures were presented to the legislators for consideration [10]. The value, which we as a nation place on learning and the acquisition of knowledge, is reflected in our support of two American institutions: the public library and a system of free, compulsory education. Two generations ago the public library became a great equalizer of educational opportunities. For our next generation the digital library can be the total equalizer. The public school system alone cannot do it, as long as there are spatial and temporal boundaries that separate opportunity from achievement. Technology can help conquer those boundaries.

The cost of acquiring content that is relevant, useful, and motivational, is too high for any individual school. Access to an information infrastructure that provides these services are the solution, but some methodology has to exist for insuring that students and teachers participate with equity in the information society. Equity of access could be organized around the concept of Electronic Serving Institutions. These institutions would orient content towards classroom and homes. Private forces will build the connectivity infrastructure, but driven by market forces it may leave out our schools and libraries. The concept of a universal infrastructure is much more specific to the needs of our learning environments, because that concept includes the inner cities and rural areas. It took half a century to electrify rural America and specific legislation called The Rural Electrification Act to reach a similar goal once before in our history [11]. Knowing this, having had that experience, we should now include that consideration and go after a more inclusive plan.

But equity will not be achieved through a universal infrastructure alone. More than connectivity is required. Assuming that equitable distribution of content is made through electronic serving institutions, and that the infrastructure reaches all the schools, there still remains a vast discrepancy in the distribution of wealth needed to buy the user-end equipment. Consider that, while it would be unthinkable to bring out of circulation, discarded, and obsolete old buses back to work in transporting our students to and from school, while no-one would think of asking for stale bread for the lunchrooms of our schools, we nevertheless, routinely request and have to depend on the generosity of government and industry for old computers to bring to our less privileged classrooms. We need to think again, and ask not for what is affordable and available, but for the best that technology has to offer our schools. Equity funding for computer equipment is as important an issue as equity access.

Finally training is the key, for teachers and parents. They cannot be left out of the information society, because if they are, a cycle will be perpetuated, and the generation gap of technology acceptance will exist forever. Programs must be instituted to provide adequate training for teachers and parents in order to be able to up with ever evolving technology as a nation. "The days when No.2 pencils and chalkboards were all the supplies teachers needed have gone," [12], was the simple, perhaps redundant statement by U.S. Education Secretary Richard W. Riley. A statement not well understood yet.

Evolution of the Infrastructure

The process of evolving our current information infrastructures (publishing, libraries, the media), and our current communications infrastructures (telephone, cable, satellite) into a coherent system called the NII (National Information Infrastructure) or the GII (Global Information Infrastructure), will require a series of evolutionary changes. These changes will consistently supply more information to more people and for many purposes. Discretely speaking, this evolution to digital information sources will start with the conversion of sources of content into CD-ROMs, then to file servers for store and forward access, and eventually to real-time direct access servers over broadband networks. Projects must be designed to contribute to the acquisition of digital content in such a way that the labor-intensive aspect of building the NII/GII (the transformation of content to digital form) is done with digital preservation in mind. The challenge is to represent and organize the content so that it can be used effectively and widely in many domains through all the phases of access evolution, and remain viable through these phases.

With respect to the ability to access this content there are problems and trade-offs particular to each phase. CD-ROMs, like books, videocassettes, and laser discs suffer from the same deficiency, lack of an automated management infrastructure. Only digital libraries will be able to tap into common layers of software that manage and organize information in the background. The problem of information management is a serious inhibitor for the harnessing of information, and so is availability. Observers of the CD-ROM market indicate that deciding on quality is also a problem, particularly in the evaluation of educational merit; and finding those CD-ROMs that are known to be of good educational value is also hard due to market demand [11]. The main advantage of the technology lies in its usefulness independent from the communications component.

With respect to the store and forward approaches, consider for instance the building of the early factories of the Industrial Age. These housed power generators because there was no reliable infrastructure that could supply remotely, the power needed to run these factories. That can be made as an analog to the current state-of-the-art in digital libraries. Eventually, as superconnectivity increases it will be possible to access servers remotely and tap into their sources of information without having to keep local caches of information. The need for resources to cache information for even viewing it, tremendously inhibits productivity. In the final stages of transformation an infrastructure will emerge that will provide the ubiquitous access that is desired, while still preserving the role of the many information servers that are now in operation. This web of information servers will probably go through its own series of transformations becoming more powerful and also more distributed. The overwhelming and somewhat discouraging observation is that the information revolution will not be revolutionary at all, but mostly evolutionary in nature.

Conclusion

There is much to be done, and the work is characteristically collective. One important thing to keep in mind is that the technologies will mature, and the infrastructures will be in place, much sooner than the time required to understand how to participate in all this. There is fear that participation may be dictated as opposed to chosen, and that benefits may accrue only to a chosen few. But there is no reason to believe or accept that. Enough work is being done and enough doors are opened. The efforts described here show that. Parents,

teachers, and students can be at these tables. We can argue convincingly that they need to be, but they also have to want to. Therefore the first order of business is to inform parents, teachers, and students so that they will want to participate.

What have been described in this document are avenues for participation, anchored in real-life experiences. These are models for the work to be done, and they are, in each case, replicable models. We need more forums for exploring the issues, more demonstration projects, and more legislative efforts. That is the work ahead. There is every reason to be hopeful, there is excitement, and there is expectation, all of the ingredients for success. If we combine that with know-how, which abounds in this country, we cannot but win.

References

1. Vision for a 21st Century Information Infrastructure, Council on Competitiveness, Publications Office, Washington, D.C., US, December 1993.
2. Masullo, M., Huang, K.T., "The EduPort Demonstration Project", Proceedings of the East-West Conference on Computer Technologies in Education, Crimean, Ukraine, September 1994.
3. Technology Panel: Clement J., Gomez L., Masullo M., McClintock R., Nguyen T., Richards J., Van Vliet D., "Workshop on the Role of Digital Libraries in K-12 Education", Conference on Information and Knowledge Management, Washington, D.C., US, November 1993.
4. Applications Panel: Brett G., Calabrese A., Hunt L., Jones B.F., Snow A., Webb L., "Workshop on the Role of Digital Libraries in K-12 Education", Conference on Information and Knowledge Management, Washington, D.C., US, November 1993.
5. Policy Panel: Gallegher G., Halem M., Kaske N., Ramirez R., Withrow F., "Workshop on the Role of Digital Libraries in K-12 Education", Conference on Information and Knowledge Management, Washington, D.C., US, November 1993.
6. Chu, W., "An Approach to Systems Management Research: Behavior Characterization", IEEE Proceedings of the First International Workshop on Systems Management, University of California at Los Angeles, Los Angeles, California, US, April 1993.
7. Masullo, M., Calo, S., Nguyen, T., Willner, B., "Multimedia On-Demand and the Organization of Education Systems", Proceedings of IFIP WG 3.2 Conference on University Uses of Visualization, University of California at Irvine, Irvine, California, US, Elsevier Science Publishers B.V., July 1993.
8. "A National Information Network: Changing Our Lives in the 21st Century", Review of the Institute for Information Studies, Edge Publishing, Washington, D.C., US, 1992.
9. Competition Policy: Unlocking the National Information Infrastructure, Council on Competitiveness, Publications Office, Washington, D.C., US, December 1993.
10. Statement of U.S. Senator Bob Kerrey (D-Neb) Before the U.S. Senate Commerce, Science and Transportation Committee, United States Senate, Washington, D.C., US, May 25, 1994.
11. Rural Electrification Act: Work of the late U.S. Senator George W. Norris (R-Neb), United States Senate, Washington, D.C., US, May 1935.
12. Jordan, M., "A Toll-Free Data Road?", The Washington Post, Washington, D.C., US, May 26, 1994.
13. Mossberg, W., "So Many CD-ROMs, So Few Ways To Find The Ones Your Want", The Wall Street Journal, New York, New York, US, December 29, 1994.